DOCUMENT RESUME

ED 272 571 TM 860 487

AUTHOR Walberg, Herbert J.
TITLE National Statistics to Improve Educational Productivity.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE [Oct 85]
NOTE 27p.; In: Invited Papers: Elementary/Secondary Education Data Redesign Project, October 1985; see TM 860 450.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Academic Standards; Adaptive Testing; Computer Assisted Testing; Computer Oriented Programs; *Data Collection; Elementary Secondary Education; National Surveys; *Productivity; Research Design; *Research Methodology; *Research Needs; *Statistical Surveys
IDENTIFIERS National Assessment of Educational Progress; National Bureau of Standards

ABSTRACT

The value of statistical research depends on valid comparisons which can usefully influence educational policy. Educational research needs to extend the measures of learning (such as the National Assessment of Educational Progress) through nationally-calibrated absolute measures and through computer-assisted and adaptive testing. Direct sampling by computer would make educational polls and national assessment faster and less expensive. A national bureau of educational standards might be founded to collect, coordinate, calibrate, archive, analyze, synthesize, and make available the data that is needed to improve educational productivity. Nine factors promoting efficiency and productivity of learning are classified in three broad categories: (1) student aptitude, including ability, development, and motivation; (2) instruction, including amount of time students engage in learning and quality of instructional experience; and (3) environmental factors, such as home climate, classroom social group, peer group, and television viewing. The tables in the appendix illustrate the effect of productivity factors on achievement revealed by quantitative syntheses carried out by a number of investigators in Australia, Canada and the United States during the past decade. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED272571

National Statistics to Improve Educational Productivity

Herbert J. Walberg

University of Illinois at Chicago

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. Manchames

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 860 487

National Statistics to Improve Educational Productivity

Herbert J. Walberg

University of Illinois at Chicago

Now is the time for the U.S. Department of Education to expand and improve national, state, and local statistics collected on education. With justification, many national reports, including A Nation at Risk (U. S. Commission on Excellence in Education, 1983), urge educational reforms; and state legislators and other groups are beginning substantial changes in educational policies. Not only is it in our national interest to study the effectiveness and efficiency of education, but the magnitude and effects of current reforms deserve the closest scrutiny during the next decade.

The government role in education may be more complicated in the U.S. than in other countries; the unique qualities of our system may both help and hinder the effectiveness of education and the collection and use of statistics. Unlike most other countries, we have no centralized governing and data-gathering ministery of education; (education policy is largely a state and local responsibility). Nor do we have, like many other advanced countries, a single centralized national agency to collect and analyze statistics on education, business, health, agricultural, cultural, and other matters.

It is not my purpose here to take sides on complicated questions of federal, state, and local control of education; nor, for that matter, on the proper influence of school administrators, teachers, parents, and students on educational goals and means. It should be recognized, however, that the the U.S. Office of Education (predecessor of the Department of Education) was created by Congress earlier in the century to collect statistics with a view toward improving education. Today, moreover, special interest and political groups ranging from liberal to conservative agree that valid data are required to formulate effective educational policies (Cooke, Ginsburg, and Smith, 1985).

Meter Sticks for Education

The value of of statistical research depends on valid comparisons, several of which can usefully influence educational policy. These include comparisons of the U.S. with other countries; among the 50 states; and among students, classes, schools, districts, and regions. In principle, all of these may be "cross-sectionally" compared at a single point in time, or changes in them may be "longitudinally" compared across years.

Even though education spending is one of the larger and growing fractions of the gross national product, and even though investments in "human capital" may be vital to future welfare (Walberg, 1983, 1984), education statistics are fundamentally invalid in several ways. The most important flaw is the lack of universalistic, absolute measures of the primary goal of education--learning.

Since Alfred Binet's turn-of-the century precedent, test developers have normatively compared students to one another rather than to absolute standards of performance. They have developed tests to reveal differences within homogeneous groups of students by selecting items that about half

the students can correctly answer and which yield "normal," bell-shaped distributions (or positively-skewed distributions from easier "mastery" test intended to detect slower learners for remediation). Whether teacher-made or commercially-developed, such tests may be administered to intact classes or even larger groups to measure learning fairly accurately but over a narrow range.

A typical range of correct answers on a 40-item test might be from 20 to 40 points correct, or a ratio of two to one between highest and lowest scores. But some students might have guessed the correct answers on a fourth of the questions and deserved lower scores; and others might have been able to demonstrate more impressive knowledge had the test been longer or composed of more difficult items suited to their ability.

Such tests are reasonable if the purpose to compare students within a narrow group with one another. Researchers might, for example, compare students within a grade level, or a school district or state on standardized, commercial tests. Similarly, teachers ordinarily compare students on material covered in their lessons: Their students may be ranked relative to each other; or a percentage correct score may be calculated. Such practices are ingrained among educators and test developers; and it may be be difficult to think that learning and other accomplishments could be distributed other than a normal curve in a narrow range. A little reflection, however, about the following question and tasks illustrates the contrary.

What distribution will result from the following tasks? Name the streets of Chicago. Throw a discus as far as you can. Prove mathematical theorems. Name the presidents of the U.S. in order. Name the capitals of Asian nations. Write Urdu verbs and French nouns. Play as many tunes as you can on the violin. Give the rules of National Collegiate Debate. Show how much weight you can bench press. Demonstrate chess openings. Put your pocket money on the table. Run as far as you can at the rate of four minutes per mile. Count the number of times you have been in trouble with the police. Enumerate the articles you have written in the school newspaper. List the countries you have visited. Repeat the numbers read to you.

Obviously, few students could approach maximum human performance on these tasks. Indeed, many would attain a score of zero on most tasks, although a few would far exceed the mean perhaps by a multiple of five, ten, or more. Human performance varies considerably more than the usual norm-referenced tests can reveal.

Absolute Measures

These distributions are of measures or counts of instances of what might be called absolute measures on ratio scales. They have a definite zero point and can range over magnitudes, some indefinitely. They are similar to the absolute measures such as meters and kilograms in the natural sciences, time and money in economics, counts and measures of behavior and perception in experimental psychology, and scales of ordinary experience. Educational research, in comparing individuals to means, percentiles, and other relativistic norms of groups, has denied itself absolute measures that have made for fundamental understandings and comparable empirical results of the natural sciences, and possibly great increases in productivity such as those in agriculture and industry.

The eminent psychometrist John B. Carroll (1982) finds little fundamental progress in mental testing since Thurstone in 1925 (pp. 67, 77). Had psychometrists continued Thurstone's efforts to calibrate mental

abilities and tasks to chronological age and learning time, however, educational productivity might be much better understood and optimized today, a half century later. Before research can establish the magnitude of the dependence of academic learning on its causal factors rather than simply the sign of the dependence, it may be necessary to develop adaptive, cardinal measures of learning and its factors over many magnitudes. Such a development, following productivity research in agriculture, health, and industry, might lead to similarly great strides in enlarging human achievement and accomplishments.

Oddly, the models that might be followed to develop absolute educational measures are outside the mainstream of academic measurement. Athletic coaches and fans undoubtedly have the most accurate and wide-ranging statistics; they can compare speeds, weights, ratios of wins and losses, and other absolute indicators. They can, for example, easily compare the times of a 100-meter dash of people of different ages, countries, and years. So also can someone who has devoted little time to studying sprints times; comparisons and their meanings can be readily understood by non-specialists.

New training regimens in athletics can be evaluated and individual performance over time can be assessed according to absolute measures. It seems that the century's steady progress in athletics--for example, the Olympics--is in part attributable to well-defined, absolute scales accurate to many significant digits across wide variations in performance. Typing tests of speed corrected for accuracy offer another useful precedent.

What about academic performance? What is lacking are absolute standards or measures that would enable us to compare children of different ages, grade levels, and abilities. It is though each test publisher and teacher had a different meter stick; and yet there is no way to equate them. Thus, test scores in a California district can not be validly compared with a New York district unless the same test happened to be administered in both districts. If a single district switched tests, it would be unable to compare earlier and subsequent scores.

If the test publisher changed editions or "re-normed" the test, then comparisons of earlier and subsequent scores are likely to be invalid (even the Scholastic Aptitude Test appears to have drifted in difficulty over the last few decades). Similarly, we cannot compare the performance of third and sixth grade students to find out how much they have learned because comparisons are strictly valid for only students who have taken the same test (not the forms for earlier and later grade levels); and the tests are capable of measuring only a narrow range of a few grade level equivalents.

One possible solution is to calibrate items and tests to national standard tests. The National Assessment of Educational Progress may offer a reasonable basis of national standards; but it represents only three age-levels, 9-, 13-, and 17-year olds. Expanding the National Assessment and coordinating it with other large-scale testing programs could lead to a more accurate picture of U.S. achievement and the possibility of a universal "meter-sticks" of learning to which other measurements may be calibrated. Given calibration formulas and procedures, states and school district staff, citizens, parents, and students could compare their scores with their previous performance or with the progress of others.

Computer-Assisted and Adaptive Testing

Even NAEP-based calibration, however, would adhere to this century's convention of giving each child within a class or grade the same test, under what would be called "batch processing" in industry. A far more efficient

and time-saving approach is "tailored-testing" (see Carroll, 1982) which flexibly adapts test items to students over great ranges of ability (rather than the reverse).

For several decades, it has been possible both in principle and in practice to program computers to assign the most discriminating items to each student, based upon her or his prior responses during the testing session (Carroll, 1982). As few as 15 tailored items can yield scores as reliable as 90 batched items suited to the average student. Alternatively, 90 tailored items given in two hours can yield very accurate assessments not in one subject but in all the major subjects of the standard curriculum. Or, 90 items could provide highly detailed assessments of skills in a single discipline, for example, word choice, grammar, spelling, and punctuation in written composition.

The increased efficiency in time use and the computer's capacity to record large amounts of information make it feasible to monitor individual student progress more frequently, accurately, and comprehensively. With a thorough, continuing assessment of what each student need to learn, it should become equally feasible to provide computer-adapted or tailored instruction. Such instruction is by no means a panacea, but it is among those educational methods that provide moderately superior achievement; and it has the further advantage of saving students' study time (Walberg, 1984). It can be expected that hardware costs will continue to fall, while software increases in sophistication and interest.

There is no reason why schools alone should provide computer-based assessment and instruction. It might be argued that since the schools have changed their basic technology of explanation, recitation, and seatwork very little since the turn of the century, other agencies might also be given an opportunity to explore these new opportunities.

In principle, students and parents could monitor student progress on absolute scales provided by entrepreneurial public and competing private groups. At a current, one-time cost of a few hundred dollars for a "dumb terminal" (without programs and memory) and a modem (to convert telephone-acoustic and computer-electronic signals), they could call a large, "mainframe" computer from their homes an 800 or a local number, take a tailored test on any subject, and bill the cost of a few dollars per assessment to a private credit card.

Public and private schools, state departments of education, and proprietary corporations could provide not only assessments but instruction as well in this way. A state and local community could finance such a system by providing an education credit card worth, say, $3,000 per year to be spent on educational services parents and students chose. These might include a mix of home and school computer instruction as well as conventional school instruction. State and local educators could suggest minimum competencies and hours of study, require performance levels for passing from one grade to the next and graduation, or impose a great number of regulations and certification practices.

Conventional and new services might range from traditional instruction in neighborhood public schools to computer-based education offered by in-state and out-of-state public and private schools and for-profit corporations. The services might be provided in schools, shopping centers, mom-and-pop neighborhood outlets, or in homes at any hour of the day or time of the year as the need or interest arose.

It is possible to program computers to monitor student progress in relation to activities on terminals and in other educational experiences. Automated statistical analyses can show which activities lead to the highest

rates of learning. Nationally-calibrated absolute measures would enable states, local districts, teachers, parents, and students to weigh the costs and benefits of various activities in making their educational plans and choices.

Some may find such new forms of instruction very much to their liking; others may find them undesirable; that is human nature. Adaptive testing on absolute scales, however, can be made carried out separately from the means of instruction. It can provide convenient measurements to assess students and programs when they seem needed.

National Statistics by Computer

Wassily Leontief, Nobel-laureate economist and inventor of national input-analysis finds it most difficult to apply his methods here in the United States where he first developed them: "The United States is the only advanced country in the world that does not have a central statistical office. Each department of the government collects statistics in the area of its own particular interests. Users of such data spend much of their time trying to reconcile and align information coming from these different sources" (Leontief, Duchin, and Szyld, 1985, p. 419).

Yet, even within the Department of Education, statistics are uncalibrated, unsystematically collected and archived, and poorly analyzed to guide national and local educational policies. Another country that keeps close track of national progress offers an interesting example of what can be done.

Growing faster economically than Japan is Singapore, where physical resources are scarce and "human capital" is taken very seriously by Prime Minister Lee Kuan Yew and others in the central government. Officials in the national ministery of education can call up in an instant any student's or young adult's test records or a mass of them for comparison from national computer bases. They can find the most qualified person to fill a particular job, or tell an American entrepreneur or manufacturer the number of trained people available for various high-value-added, growth industries such as tourism, electronics, and petroleum processing.

It may be argued that the U.S. may be too big, diverse, complicated, and perhaps fixed in its educational ways to enact such an innovation. Besides, possible abridgment of freedom and confidentiality are likely to worry and deter educators from something so intrusive. Still, the Internal Revenue Service does not seem to have abused its vast powers to reveal confidential information; if anything, there seems more need to worry about the accuracy of the information supplied by taxpayers. There seems no good reason to think that measurement calibration and related services provided by the government would violate privacy.

Citizens, moreover, particularly poor and minority-group parents, seem more enthusiastic than the educational establishment for better measurements and higher standards. Conducted for business, civic, and minority organizations, for example, a recent poll of 1,816 Chicago residents showed 88 percent feel that all high school students in the U.S. should be be required to pass a standard examination before being graduated.

In addition, Chicagoans overwhelmingly favored a tougher curriculum for high schools: The lowest-income groups most favored extending the curriculum to more subjects; and blacks more often than others preferred the toughest requirements in science, history, and foreign languages for college- and non-college-bound students. Of Chicago adults with children in the public schools, 69 percent said they would send their children to private schools if they could afford it primarily because they would get a better teaching, attention, and discipline (Walberg and Hess, 1985).

If we start from the premises that we must inform citizens about their schools; that educators should be informed about their business including their costs, benefits, and views of citizens; and that better education statistics may help us to understand and solve our educational problems--then we need to think about harnessing the vast powers of the computer, as other industries have done, to increase competitiveness and productivity. School districts and state departments of education are enlisting computers in central offices and classrooms; and it would seem the proper role of the federal government to lend research support and technical assistance to help coordinate the efforts. Comparisons of all sorts will be more valid, other things being equal, to the extent that data are obtained uniformly.

In addition, national hook-ups, perhaps sponsored by the federal government, would make it feasible to conduct sample surveys of districts, schools, and students directly by computers. Students, for example, could rapidly complete tailored tests and questionnaires by terminal and modem. In compensation, they might be offered a small stipend, or at least they and their schools could receive an immediate summary of results which could also be provided by long-distance telephone connections to state and federal computers.

The further advantage is the speed at which such surveys and tests can be completed. The time-consuming steps of printing tests and questionnaires, mailing, key boarding and screening data, and the like can be skipped. Even analyses can be automated.

Like Gallop and other polls of 1,500 respondents that provide reasonably accurate estimates of public opinion in the nation, direct sampling by computer would make educational polls and national assessments fast and cheap; they would minimize the total human time answering questions yet provide more accurate estimates than far larger but unscientific surveys. Quarterly or even monthly survey reports on important measures could be made routine as they are in commerce and industry. Local, state or national assessments of special topics might be commissioned and completed in less than a month. In principle, we would not have to wait a year for the Phi Delta Kappan's Gallop Poll on education, several years for cycles of the National Assessment nor as much as a decade between International Studies.

A National Bureau of Educational Standards

Before turning to the kinds of measures that seem desirable to collect, it should be acknowledged that what is called for above is a tall order as compared with what has been planned and spent on educational research. What is spent on educational research by other standards, however, is minuscule. Even if spending on educational research amounted to $150 million annually (Walberg, 1983), it would be less than .006 percent of annual educational spending on public elementary and secondary education in recent years. By comparison, it is by no means unusual for growing corporations in competitive industries to spend 5 or 10 percent of annual revenues on research and development.

The costs of federal research on defense, space, and medicine obviously dwarf expenditures on educational research which may pay greater dividends for the nation's future welfare. As Adam Smith said and Japan demonstrates, human capital is just as important as physical and financial capital in determining the wealth of nations. And it is clear that education can be made much more productive in increasing the ratio of its benefits to costs.

To sustain the coherent programmatic data collection that seems required may require a agency of the U. S. Department of Education. Such

an agency, perhaps called the "National Bureau of Educational Standards," would be analogous to that first created in Paris long ago for keeping standards of weights and lengths. Such an agency would have to be carefully planned and under close scientific scrutiny since it would have to provide precise definitions and measurements of education and learning, which tend toward imprecision, non-comparability, and intractability.

Such an agency would need to avoid partisan stances, value judgments, and declarations of what constitutes adequacy or excellence. Like the National Bureau of Standards which provides physical standards for our country, it would have to adhere to scientific and factual questions rather than values stances, inasmuch as it is possible in education.

It is the charge of the Department of Education to collect statistics. It appears, moreover, that no other agency, public or private, could take on the large task of thinking through, commissioning, and monitoring or conducting the research required to put such an agency in place.

Aside from calibration, a National Bureau of Educational Standards could serve as the central government repository and publisher of statistics on education in the U.S and, where appropriate and feasible, in other countries. In addition, for those who wish to analyze the raw data rather than examine pre-digested summaries, a National Bureau could serve, like the Library of Congress, as an archive of computer tapes of educational data that could be reproduced at cost by requests in writing, in person, or by telephone (including telephone requests for data transfers by computer).

It would have a capable technical staff to archive data in standard formats that could reproduced for secondary analysis by investigators in universities, state departments of education, schools, and newsrooms. In this sense, it could be modeled after the Institutional Consortium for Political and Social Research at the University of Michigan that archives and makes available major social surveys and public opinion polls. Although such surveys may costs hundreds of thousands if not millions of dollars, the tapes and codebooks for any can be supplied at cost a few hundred dollars. With artificial intelligence, it should become possible within a decade or two for non-technical people to query such data bases for a few dollars by voice over the telephone without having mastered programming.

The beginnings of these functions are already represented in the National Center for Educational Statistics that distributes the High School and Beyond data on about 58,000 sophomores and seniors, their parents, and teachers. In addition, the Educational Commission of the States made available and Educational Testing Service currently makes available at cost data on several-hundred thousand 9-, 13-, 17-year-olds and young adults collected in the National Assessment of Educational Progress. In addition, the tapes from the many surveys of nearly 50 countries participating in the International Association for the Evaluation of Educational Achievement could easily be copied and archived, as could other surveys such as the Gallop Polls on education, the Equality of Educational Opportunity survey, the National Longitudinal Survey, and the General Social Survey. These data sets were assembled during at a cost of perhaps $500 million and are largely under-analyzed. A National Bureau could serve as archive, calibrator, synthesizer, reporter, and at-cost distributor or raw data and results.

Current Federal Statistics

It has been said that democracy is the worst government except for all other forms. The same may be true of current statistics the U.S. Department provides. Therefore, we should be loath to stop collecting any data series in education that has already been started, even on things that

have seemed of little bearing on learning. If anything, we must expand the collection and coordination of data, and encourage scholars and others to analyze the costly and valuable statistics that are currently available.

Federal government spending on education statistics, however, is small by several standards. In school year 1982-83, for example, spending on public elementary and secondary schools in the U.S. by federal, state, and local government was respectively $56, 52, and 8 billion, which comes to a total of $116 billion, which is 4.5 percent of the $2.6 national income (Indicators, 1985, p. 22). If the federal government spent $100 million on better educational statistics, it would be amount to less than one-tenth of one percent of total educational spending on public schools and might increase efficiency by many billion. Given U.S. government spending of $1.4 billion on statistics (Alonso and Starr, 1985, 123), education's 4.5 percent share (based on the public school percent of national income) would be $63 million, in contrast to $8.7 million in current spending by the National Center for Educational Statistics. Higher spending should yield better statistics and make the "education industry" more comparable to agriculture, medicine, and various industries that base practice upon productivity comparisons.

Even the aggregate and crude numbers on costs and enrollments now compiled by the federal government from data supplied by the states can raise pointed questions. For example, the 1985 Indicators published by the U.S. Department of Education shows that an average of $2,948 was spent on each of the 39.6 million children in public schools in 1982-83 (p. 22). (By comparison, according to Feistritzer, 1985, the average per-student costs of Catholic schools in 1982-83 was $782; and private school tuition was $1,029.)

Public school teachers reported an average of 24 students in their classes; and the average ratio of students to all full-time equivalent teaching staff (which includes special teachers) was 19. So, depending on the estimate, per-student annual spending was $55 or $70 thousand per teacher (p. 30). Teacher earnings, however, were about $19 thousand (Indicators, 1985, p. 30), or only about a quarter or a third of total costs (actually the true fraction must be even smaller, since complete earnings which includes moonlighting and summer work are included as earnings).

Thus, indirect costs in public schools are apparently two or three times as high as the direct educational services provided by teachers. Where is this extraordinary amount of money going if not to teachers? Does it account for the approximate 500 percent increase in inflation-adjusted, per student costs since 1930? Has academic achievement gone up accordingly? Is what is being provided by indirect costs as valuable to students as teaching services? Does it help teachers to do their jobs more efficiently? Can it be going to physical facilities during a period of declining enrollments? If it is going to administrators, can such heavy bureaucratic spending be justified? Are federal and state governments creating local bureaucracies to deal with special programs and complex regulations? Do any of these explanations fit with corporate trends toward lean, competitive organizations with the most senior administrators close to customers rather than layered away from them by corpulent staffs?

Whatever the answers to these questions, the numbers themselves are provocative; they stimulate discussion and research. Such data should be easily accessible so that the public and educators can deliberate about them. It is important to keep accurate tabulations of them over the years so that we can better understand how the levels of learning are changing, what is and is not changing them, and what might be done to increase effectiveness. We should be reluctant to omit any just as the Library of Congress avoids

dropping subscriptions of unfashionable magazines.

The Quality of Federal Statistics

Even though we should continue and expand the collection of statistical series, we must be duly cautious and consider, as Aristotle advised, the source. The British statistician Sir Josiah Charles Stamp (1880-1941) warned:

> "The government are very keen on amassing statistics. They collect them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damm pleases" (quoted in Alonso and Starr, 1985, p. 123).

Cooke, Ginsburg, and Smith (1985) compiled several alarming discrepancies in estimates in important national statistics on and related to education. School safety and security, for example, have found to be related to learning gains; but how safe are students? The National Crime Survey administered by the U. S. Department of Justice seems to indicate from household interviews that about 10 percent of junior and senior high school students are victims of assault, robbery, or theft each year. The National Institute of Education, however, reported from confidential answers by students in their classrooms that 10 percent were victimized each school month--an estimate at least ten times larger than the Department of Justice estimate based on parent interviews about their children.

Similarly, the U.S Department of Education's Vocational Education Data Systems reported 741 thousand New Jersey students taking high school vocational education courses in 1979--a number that exceeded the State's high school enrollment by more than 50 percent. In Virginia, the 29 thousand Indians which VEDS indicates as enrolled in vocational education represents more than three times the total Indian population, according to the State Indian Commission.

Even well researched variables in the mainstream of educational reform movement are suspect. Instructional time in the United States is rarely more than 60 percent of the school day; but the share varies by more than 2 to 1 among schools, and engaged time is only a varying fraction of allocated time, according to Cooke, Ginsburg, and Smith (1985). Japan's high schools may employ twice as much engaged time in the four years of high school (including extramural study) to yield achievement equivalent to the U. S. bacchalaureate in mathematics, science, geography, native and foreign languages, and music as well as non-academic pursuits (Walberg, Paschal, and Weinstein, 1985). It is, however, difficult to get more than ball park estimates of these important comparisons.

In Illinois, perhaps because of inflation of course titles or blurring of content, 80 percent of the high school students reported taking geometry, but a census of actual transcripts in the State revealed that only a quarter had. In California, 99 percent school attendance is reported; but students who have "valid" excuses there are reported as in school--a far different definition is given in other states (Cooke, Ginsburg, and Smith, 1985).

It may be hoped that different means of reporting, biases of the reporters, and random factors may balance out and permit at least rough comparisons across respondents, states, nations, and time periods. But it remains a vague and often patently false hope; and the discrepant estimates of Cooke, Ginsburg, and Smith (1985) may chime the thirteenth hour on the educational statistics clock for some important figures. "Lies, damn lies, and statistics" said Benjamin Disraeli and Mark Twain.

11

The National Bureau Revisited

Science offers several ways to assess and solve such statistical problems; they deserve the support of a National Bureau of Educational Standards. One is to insist that highly detailed, explicit, and publicly accessible descriptions of data definition and collection procedures. Another is to commission papers and convene conferences to criticize, design, and redesign large national and international sample surveys. To some extent, these are major functions of refereed journals and associations in the natural sciences and their applications. They are often funded by U. S. government agencies such as the Department of Agriculture, Department of Health and Human Services, and the National Science Foundation in the such cases of expensive health and productivity surveys and massive projects in physics.

National groups of blue-ribbon layman or practicing professionals have tried to carry out these functions in preparing recent national reports on education. They have employed school visits, hearings, and a selection of expert testimony and papers. Democratic societies should allow, indeed, encourage non-technical deliberation and formulation of policy. Public and private commissions should deliberate and recommend goals, values, and means, which may be enacted by legislators, private agencies, and individuals.

But such groups should have accurate statistics as one basis of their deliberations. They may not have the technical competence to gather and assess the statistical facts; and technical experts may not be able to see beyond the facts to the public interest. In view of the limits of human time, skills, and knowledge, some specialization of function is required. A National Bureau of Educational Standards should be restricted to collection and assessment of data, calibrating and correlating measures, commissioning large-scale studies, making information available, and criticizing it. In this way, it may provide good data for policy analysts and decision makers.

It should, however, avoid political stances and recommending of policies and practices. The National Bureau of Standards accurately tells us how long yards and meters are, not how long our houses or apartments should be. The Department of Labor gives the incomes of occupational groups (with a margin of error) not evaluations of what income distributions should be; when it gets beyond the ascertainable facts to such predictions as the number of mechanical engineers required in ten years, it is often wrong. The Department of Agriculture can give the average corn yield of Iowa farmland and the increments associated with degrees of tillage, irrigation, and fertilizer; but the farmers decide how to farm. These seem instructive precedents.

What Educational Data Needs Collecting?

Following the lead of early agricultural experimentation, much educational research focuses on the relation of single causes and effects. Education, however, obviously involves many means and ends, each with an explicit or implicit cost or value. The promotion of efficiency requires the specification and measurement of the chief causes, means, or "factors" of production.

Experiments and statistical studies of productivity data together with cost and value estimates have enabled a wide variety of industries to increase the value of their output while simultaneously reducing costs thereby raising human welfare. Although such thinking may seem alien to some educators, the public ranks research on educational effectiveness

higher in priority than most other fields of scientific investigation in the natural and social sciences (Walberg, 1983); and educators may do well to think more explicitly and unsentimentally about our business and to try to found it on the emerging consensus of scientific evidence.

It should also be said, however, that we educators are far from estimating explicit costs and values. The prior problem, now being solved, is estimating the magnitudes of effects of educational inputs on outputs, which primarily involves causal rather than value questions. It is these chief causes and effects that deserve first priority in national data archives.

Nine factors require optimization to increase affective, behavioral, and cognitive learning (see Walberg, 1984, and the cited references for a more detailed discussion). Potent, consistent, and widely generalizable, these nine factors fall into three groups:

Student aptitude includes:

1) Ability or prior achievement as measured by the usual standardized tests,
2) Development as indexed by chronological age or stage of maturation, and
3) Motivation or self concept as indicated by personality tests or the student's willingness to persevere intensively on learning tasks.

Instruction includes:

4) the amount of time students engage in learning and
5) the quality of the instructional experience including psychological and curricular aspects.

Four environmental factors also consistently affect learning:
the educationally-stimulating, psychological climates of
6) the home,
7) classroom social group, and 8) the peer group outside school; and
9) minimal leisure-time television viewing.

The first five aspects of student aptitude and instruction are prominent in the educational models of Benjamin S. Bloom, John B. Carroll, Robert Glaser, and others. Each appears necessary for learning in school; without at least a small amount of each, the student can learn little. Large amounts of instruction and high degrees of ability, for example, may count for little if students are unmotivated or instruction is unsuitable.

These five essential factors, however, are only partly alterable by educators since, for example, the curriculum in terms of lengths of time devoted to various subjects and activities is partly determined by diverse economic, political, and social forces. Ability and motivation, moreover, are influenced by parents, by prior learning, and the students themselves. Thus educators are unlikely to raise achievement substantially by their own efforts alone.

Of the remaining factors--the psychological climate of the classroom group; enduring affection and academic stimulation from adults at home; and an out-of-school peer group with learning interests, goals, and activities--influence learning in two ways: Students learn from them directly; and these factors indirectly benefit learning by raising student ability, motivation, and responsiveness to instruction. In addition, about ten (not the more typical 30) weekly hours of television viewing seem optimal for learning, perhaps because more television time displaces homework and other educationally-constructive activities outside school.

The major causal influences flow from aptitudes, instruction, and the psychological environment to learning. In addition, however, these factors also influence one another, and are also influenced in turn by how much

students learn, since those who begin well learn faster.

Other social factors influence learning in school but are less directly linked to academic learning. For example, class size, financial expenditures per student, and private governance (independent or sectarian in contrast to public control of schools) weakly correlate with learning, especially if the initial abilities of students are considered. Thus, improvements in the more direct and more alterable factors hold the best hope for increasing educational productivity.

Thus, in my view, school and district economic, political, and sociological characteristics and conditions are less relevant to learning because their influences are less alterable, direct, and observable. They are not substitutes for the nine factors, but more distant forces that can support or interfere with them.

More and less productive classes, moreover, may be expected in the same school; and it is somewhat misleading to characterize a whole school or district as effective--just as it is less accurate to characterize an optimal condition of plant growth as the average annual rate of rainfall in a state or farm than the amount of rain and irrigation that reaches the roots of a single plant in a given time period.

The educational productivity theory itself is admittedly over-simplified because learning is clearly affected by school and district characteristics as well as many economic, sociological and political forces at the school, community, state, and national levels. Yet these characteristics and forces--such as the sex, ethnicity, and socioeconomic status of the student, the size and expenditure levels of schools and districts, and their political and sociological organization--are less alterable in a democratic. pluralistic society; are less consistently and powerfully linked to learning; and appear to operate mainly through the nine factors in the determination of achievement. Thus, I offer our theory not as a threat to those who see the efficacy of other factors but as a friendly, collegial invitation to demonstrate their effects on the nine factors or directly on learning.

Methods of Research

Since our concern was productivity, we hoped that our own research would efficiently capitalize on previous inquiry; and, under the support of the National Institute of Education and the National Science Foundation, our team of investigators started by compiling reviews of the 1970s on the productive factors in learning. Next, quantitative syntheses of studies of productive factors were conducted; syntheses of several thousand investigations were compiled (see Walberg, 1984, for a more detailed account). Case studies of Japanese and American classes were carried out to compare educational productivity in the two countries. The productive factors were further probed for their significance in promoting learning in three large sets of statistical data on elementary and high school students--the National Assessment of Educational Progress, High School and Beyond, and the surveys of the International Association for the Evaluation of Educational Achievement.

Collectively the various studies suggest that the nine factors are powerful and consistent in influencing learning. Syntheses of about 2,575 studies suggest that these generalizable factors are the chief influences on cognitive, affective, and behavioral learning. Many aspects of these factors can be altered or influenced by educators.

The first five essential factors appear to substitute, compensate, or trade-off for one another at diminishing rates of return. Immense quantities of time, for example, may be required for a moderate amount of learning if

motivation, ability, or instructional quality is minimal. Thus, no single essential factor overwhelms the others; all appear important.

Although the other factors are consistent statistically- or experimentally-controlled correlates of academic learning, they may directly supplement as well as indirectly influence the essential classroom factors. In either case, the powerful influences of out-of-school factors especially the home environment must be considered.

For example, the 12 years of 180 6-hour days in elementary and secondary school add up to only about 13 percent of the waking, potentially-educative time during the first 18 years of life. If more of the 87 percent of the student's waking time nominally under the control of parents that is spent outside school were to be spent in academically-stimulating conditions in the home and peer group, then the total amount of the student's total learning time might be dramatically raised beyond the 13 percent of the time in conventional American schools.

For instance, the average of 28 hours a week spent viewing television by high school students might usefully be added to the mere 4 or 5 weekly hours of homework (Walberg and Shanahan, 1983). Europeans and Japanese believe homework helps learning; empirical results of American research support their belief.

The numerical results of syntheses of the effects in several thousand studies of academic learning conducted during the past half century. Interested readers and those who wish technical details may examine the findings and methods reported in the compilations of these syntheses (cited in the references in Walberg, 1984, which in turn, contain references to the original studies. (In several instances, separate estimates of correlations and effects are available for science and mathematics because the National Science Foundation awarded grants for special synthesis projects on these two subjects. The tables contain both effects and correlations, and the correlations assume a one- standard deviation rise in the independent variable.)

Sample survey items and descriptions of sets of items from High School and Beyond, the National Assessment of Educational Progress, and the International Association for the Evaluation of Educational Achievement provide national and international baselines in various years, and theyshould be considered as candidates on this ground alone. Better items and direct observations can also be developed and used.

Beyond Academic Achievement

If education proceeds by fads rather than cumulative research, it will fail to make the great advances in productivity that have characterized agriculture and industry in this century. It may be argued, however, that education is a complex subject and cannot be reduced to a few external benefits or measures of outcomes. This argument also applies to any enterprise: The desirability of an automobile cannot be reduced to numbers on its speed and power; bushels of corn per acre need to be considered in the light of percentage of protein per unit weight, predicted prices, fuel requirements, human labor, and the like.

Similarly, better nationally-calibrated measures of achievement including facts and "higher-order skills" in English, mathematics, science, civics, history, foreign languages, art, and music and what produces them would be a great accomplishment. But they would hardly suffice and may be misleading.

A synthesis of the relation of conventionally-measured educational

outcomes and adult success shows their slight association (Samson and others, 1982). Thirty-three post-1949 studies of the college and professional-school grades of liberal arts and business graduates, nurses, physicians, engineers, civil servants, teachers, and other groups show an average correlation of .155 of these educational outcomes with life-success indicators such as income; self-rated happiness; work performance and output indexes; and self-, peer-, and supervisor-ratings of occupational effectiveness. Thus, only about 2.4 percent of the variance in these indicators of adult success was predictable from grades given by professors.

These results should challenge educators and researchers to seek a balance between continuing autonomy, motivation, responsibility, and skills to learn new tasks as an individual or group member on one hand and mastery of teacher-chosen, textbook knowledge measured on conventional tests that may soon be obsolete or forgotten on the other. Researchers need to think again about how civic virtue, perseverance, will power, cooperation, entrepreneurship and the like that are no longer in the current psychological lexicon might be measured and encouraged.

One clue comes from old studies of open education, in which teachers and students negotiated contractual terms about what students would learn. Open educators tried to encourage educational outcomes that reflect teacher, parent, student, and school board goals such as cooperation, critical thinking, self reliance, constructive attitudes, life-long learning, and other objectives seldom considered by psychometrists. Raven's (1981) summary of surveys in Western countries including England and the United States, shows that, when given a choice, educators, parents, and students rank these goals above test scores and high marks.

Hedges, Giaconia, and Gage (1981) synthesized 153 studies of open education including 90 dissertations. The average effect was near zero for achievement, locus of control, self concept, and anxiety (which suggests no difference between open and control classes on these criteria); about .2 for psychological adjustment, attitude towards schools and teachers, curiosity, and general mental ability; and about a moderate .3 for cooperativeness, creativity, and independence. Thus, students in open classes do slightly or no worse in standardized achievement and slightly to moderately better on several outcomes that educators, parents, and students hold to be of great value. Thus, this recently-synthesized old research shows the value non-standard outcomes and demonstrates that conventional measures do not necessarily enhance or trade-off against unconventional accomplishments.

Another precedent for non-conventional measurement is the current effort beginning under the sponsorship of the Swedish Ministery of Education. Sweden is fortunate in having a longitudinal sample first measured in 1961 of people born on the 5th, 15th, and 25th of all months of 1948, who are now nearing 40 years of age. Harnqvist (1984) is beginning an internationally important series of studies of this sample to discover how early school and other experiences influence adult knowledge and attitudes. Of about 120 adult characteristics, 71 percent have shown significant partial correlations with amount of education, 30 percent with measured intelligence, and 20 with social background.

Other things being equal, Swedish adults with more education, for example, more often reported that their jobs provided them with new knowledge and more influence on determining their working conditions. Amount of education was positively associated with "cultural" activities such as going to theaters and concerts, and negatively correlated with "entertainment" through weekly magazines, television, and sports events. More highly educated men reported higher skills in cooking and lower skills

in repairing a car; and having better information on appealing decisions and less about seeking economic support from society. In the interviews, more educated men and women used more words, more different words, a greater percentage of words with more than 10 letters; and they required fewer interviewer interventions to complete their responses.

Harnqvist's is a pioneering longitudinal investigation. Since the effects of education may not turn up immediately, except on knowledge tests alone and less well on other measures, we need more such long-term studies that relate adult characteristics to educative experiences and activities within and outside school. We are fortunate, indeed, to have Harnqvist's contribution that shows the influence of amounts of education; but we would like to have more specific measurements of early and later accomplishments such as hobbies and prizes won, and experiences such as courses taken, homework hours, books read at leisure, family activities, trips abroad, military service, and the like. These would enable us to relate early characteristics and experience to later achievements and attitudes.

Conclusion

The present seems a time for great opportunity in educational reform and research in education. Agriculture, engineering, and medicine made great strides in improving human welfare as doubts arose about traditional, natural, and mystical practices, as the widened measurement of results intensified, as experimental findings were synthesized, and as their theoretical and practical implications were coordinated and vigorously implemented and evaluated.

Education is no less open to humanistic and scientific inquiry and no lower in priority since half the workers in modern nations are in knowledge industries, and the value of investments in people is now more apparent than ever (Walberg, 1983). Although it is possible to find fault with federal statistics on education, the last decade or two has been a period of quiet but significant accomplishments; and larger amounts of valuable data are being accumulated.

Recently the National Research Council's Committee on Indicators of Precollege Science and Mathematics Education issued a report calling for the national measurement and tracking of the many of the same productivity factors and outcomes discussed above (Raizen and Jones, 1985). The U.S. Department of Education, working with 16 education organizations, has already developed a plan for systematically collecting outcome, process, and context data and issued its first report, Indicators of Education Status and Trends (1985). This contains a series of data, presented both in tables and graphs over time, showing the course of education measures over several decades.

Both reports continue our tradition of collecting enrollment and spending data; but go beyond it in recommending (in the case of the NRC report) and displaying (in the DE report) changes in test scores, international comparisons of achievement, remedial college course enrollments, class sizes, verbal abilities of the teaching force, public-opinion ratings of schools, and state-required curriculum units. These reports give us hope that we may reach consensus on extending measures of learning and of the productive factors that bear upon it, and that a national bureau might be founded to collect, coordinate, calibrate, archive, analyze, synthesize, and make available the data that is needed to improve educational productivity.

References

Alonso, William, and Starr, Paul (1985) A Nation of Number Watchers. Wilson Quarterly, 9 (3), 93-123

Carroll, J. B. The measurement of intelligence. In R. J. Sternberg (Ed.), Handbook of human intelligence. New York: Cambridge University Press, 1982.

Cooke, Charles, Ginsburg, Allen, and Smith, Marshall (1985) The Sorry State of Education Statistics. Madison: Wisconsin Center for Educational Research.

Feistritzer, C. Emily. (1985) Cheating Our Children. Washington, DC: National Center for Educational Information.

Harnqvist, Kjell (1984) An Empirical Study of Long-Term Effects of Education. Tel Aviv, Israel: Paper presented at the First International Conference on Education in the 1990s.

Hedges, L. V., Giaconia, R. M., & Gage, N. L. Meta-analysis of the effects of open and traditional instruction. Stanford, CA.: Stanford University Program on Teaching Effectiveness, 1981.

Horn, Elizabeth, and Walberg, Herbert J. (1984) Achievement and interest as functions of quality and level of instruction. Journal of Educational Research, 77, 227-232.

Indicators of Education Status and Trends (1985) Washington, DC: U. S. Department of Education.

Leontief, Wassily, Duchin, Faye, and Szyld, Daniel B. (1985) New Approaches to Economic Analysis. Science, 112, 419-427.

Raizen, Senta A. and Jones, Lyle V. (1985) Indicators of Precollege Education in Science and Mathematics: A Preliminary Review. Washington, DC: National Academy Press.

Raven, J. The most important problem in education is to come to terms with values. Oxford Review of Education, 1981, 7, 253- 272.

Samson, G. E., Graue, M. E., Weinstein, T., & Walberg, H. J. (1984) Academic and occupational performance: A quantitative synthesis. American Educational Research Journal, 21, 311- 21.

Thurstone, L. L. The relation between learning time and length of task, Psychological Bulletin, 1930, 37, 44-53.

U. S. Commission on Excellence in Education (1983) A Nation at Risk, Washington, DC: U. S. Government Printing Office.

Walberg, Herbert J. (1984) Educational Expenditures and Student Achievement. New Orleans, Louisiana: Paper presented at the annual meetings of the American Educational Research Association.

Walberg, Herbert J. (1984) Improving the Productivity of America's Schools. Educational Leadership, 41 (8), 19-30.

Walberg, Herbert J. National Abilities and Economic Growth. New York, NY: Paper presented at the annual meetings of the American Association for the Advancement of Science.

Walberg, Herbert. J. Scientific literacy and economic productivity in international perspective, Daedalus, 1983, 112, 1-28.

Walberg, Herbert J., Harnisch, Delwyn L., and Tsai, Shiow-Ling. (in press) Mathematics Productivity in Japan and Illinois. Evaluation in Education: An International Journal, 1985.

Walberg, Herbert J., and Hess, G. Alfred (1985) Chicagoans View the Public Schools Chicago: Chicago Panel on Public School Finances.

Walberg, Herbert J., Paschal, Rosanne A., and Weinstein, Thomas (1985) Homework's Powerful Effects on Learning. Educational Leadership, 42 (7), 76-78.

Walberg, Herbert J., and Rasher, Sue Pinzur (1979) Achievement in the 50 States. In Herbert J. Walberg (Ed.), Educational Environments and Effects. Berkeley, CA.: McCutchan Publishing.

Walberg, Herbert J., and Shanahan, Timothy (1983) High school effects on individual students. Educational Researcher, 12 (7), 4-9.

Appendix

This appendix contains several tables that illustrate the magnitudes of effects of productivity factors on achievement revealed by quantitative syntheses carried out by a number of investigators in Australia, Canada, and the United States during the past decade. In addition, operational representations of the factors and sample items from re-analyses of the National Assessment of Educational Progress, High School and Beyond, and the first mathematics survey of International Association for the Evaluation of Educational Achievement are given (Horn and Walberg, 1982; Walberg and Shanahan, 1983; and Walberg, Harnisch, and Tsai, 1984).

Table 1

Influences of Aptitudes on Learning

Aptitude	Effect	Size
Ability		
IQ	.71	XXXXXXX
IQ (Science)	.48	XXXXX
Development		
Piagetian Stage	.47	XXXXX
Pia. Stage (Science)	.40	XXXX
Motivation		
Motivation	.34	XXX
Self-Concept	.18	XX

Note: The X symbols represent the sizes of the correlation coefficients in numbers of tenths.

Table 2
Instructional Quality and Time Effects on Learning

Method	Effect	Size
Reinforcement	1.17	XXXXXXXXXXXX
Acceleration	1.00	XXXXXXXXXX
Reading Training	.97	XXXXXXXXXX
Cues and Feedback	.97	XXXXXXXXXX
Science Mastery	.81	XXXXXXXX
Cooperative Programs	.76	XXXXXXXX
Reading Experiments	.60	XXXXXX
Personalized Instruc.	.57	XXXXXX
Adaptive Instruc.	.45	XXXXX
Tutoring	.40	XXXX
Individualized. Science	.35	XXXX
Higher-Order Questions	.34	XXX
Diagnostic Prescription	.33	XXX
Individualized Instruc.	.32	XXX
Individualized Math.	.32	XXX
New Science Curricula	.31	XXX
Teacher Expectation	.28	XXX
Computer-Assis. Instruc.	.24	XX
Sequenced Lessons	.24	XX
Advanced Organizers	.23	XX
New Math. Curricula	.18	XX
Inquiry Biology	.16	XX
Homogeneous Groups	.10	X
Programmed Instruc.	-.03	-.
Class Size	-.09	-X.
Mainstreaming	-.12	-X.
Instructional Time	.38	XXXX

Note: The X symbols represent the sizes of effects in tenths of standard deviations.

Table 3

Home, Peer, Class Morale and Media Effects

Method	Effect	
Graded Homework	.79	XXXXXXXX
Class Morale	.60	XXXXXX
Home Interventions	.50	XXXXX
Home Environment	.37	XXXX
Assigned Homework	.28	XXX
Socioecon. Status	.25	XXX
Peer Group	.24	XX
Television	-.05	X.

Note: The X symbols represent the sizes of effects in tenths of standard deviations or correlations.

APPENDIX NAEP

Variable Descriptions and Sample Characteristics

Operational Definition, Internal Consistency, Sample Paraphrased Items, Scoring, Percent of Sample in Each Category (when applicable).

Based on Sample of 2,294 Students.

Achievement
Fifty-five items assessing student achievement in five content categories and four cognitive-process levels. Alpha internal consistency reliability = .92.

Interests
Three self-report items probing student willingness to study mathematics not part of a classroom assignment. Alpha internal consistency reliability = .47. "How often did you work ahead in your mathematics book?" "How often did you do mathematics problems that were not assigned?" "How often did you study mathematics topics that were not in the textbook?" Coded: 3 = often, 2 = sometimes, 1 = never, blank = no response or missing.

SES
Highest amount of either parent's education. Coded:
1 = Not a high school graduate (15.6%)
2 = Graduated high school (34.8%)
3 = Post high school (44.9%)
blank = Unknown or missing (4.7%)

Traditional instruction
Two items on traditional instructional methods. Alpha internal consistency reliability = .55. "How often has each of the following been used in the courses you are taking this year?"
1. Listening to the teacher's lecture
2. Studying from textbooks

Coded: 3 = frequently, 2 = fairly often, 1 = seldom or never, blank = no response or missing.

Home environment
Twelve items on home characteristics. Alpha internal consistency reliability = .66. "Which of the following do you have in your home?"
1. Newspaper received regularly
2. Magazine received regularly
3. More than 25 books
4. Encyclopedia
5. Dictionary
6. Record player
7. Tape recorder or cassette player
8. Typewriter
9. Vacuum cleaner
10. Electric dishwasher
11. Two or more cars or trucks that run
12. English spoken most often at home

Coded: 0 = do not have, 1 = have, blank = no response or missing.

Sex
Coded: 1 = female (51.5%), 0 = male (48.5%)

Coded: 1 = white (81.8%), 0 = not white

Coded: 1 = black (12.7%), 0 = not black

Coded: 1 = Spanish heritage (4.8%), 0 = not Spanish heritage

Student-centered instruction
Two items on instructional methods emphasizing student participation. Alpha internal consistency reliability = .43. "How often has each of the following been used in the courses you are taking this year?"
1. Participations in student-centered discussions
2. Having individualized instruction in small groups or one-to-one with a teacher

Coded: 3 = frequently, 2 = fairly often, 1 = seldom or never, blank = no response or missing.

Stimulation
Four items indicating frequency of course-related activities. Alpha internal consistency reliability = .66.
"How often has each of the following been used in the courses you are taking this year?"
1. Working on a project or in a laboratory
2. Writing essays, themes, poetry, stories
3. Going on field trips
4. Library or media center assignments

Coded: 3 = frequently, 2 = fairly often, 1 = seldom or never, blank = no response or missing.

Number of math courses
Seven items indicating mathematics course taken or completed; summed to indicate number of years mathematics was studied. "Which of the following mathematics courses have you studied?"
1. General, business, or consumer mathematics (38.3% studies one year)
2. Pre-algebra (36.0%)
3. First year algebra (67.3%)
4. Geometry (43.3%)
5. Second year algebra (28.7%)
6. Trigonometry (6.7%)
7. Pre-calculus/calculus (2.2%)

Coded: 0 = not studied, .25 = studied less than one half of school year, .5 studied one half of school year, 1 = studied about 1 school year, blank = no response or missing.

Highest course
Indicates highest level mathematics course taken for at least one half of school year. Coded in order of course difficulty:
1 = General, business, or consumer mathematics (10.9%)
2 = Pre-algebra (7.2%)
3 = First year algebra (17.3%)
4 = Geometry (14.0%)
5 = Second year algebra (23.8%)
6 = Trigonometry (9.7%)
7 = Pre-calculus/calculus (4.0%)
Blank = no response or missing (13.1%)

Homework
Amount of time spent doing homework last night. Coded:
3 = more than two hours (8%)
2 = between one and two hours (18%)
1 = less than one hour (20%)
0 = did not do homework or no homework assigned (40%)
Blank = no response or missing (14%)

TV
Amount of time spent watching TV last night
8 = six or more hours (4.0%)
7 = five hours (3.6%)
6 = four hours (9.2%)
5 = three hours (11.2%)
4 = two hours (14.9%)
3 = one hour (8.0%)
2 = less than one hour (15.7%)
1 = none (18.6%)
Blank = no response or missing (14.9%)

BEST COPY AVAILABLE

TABLE I

Variable Descriptions and Numbers of Items, Reliabilities, and Univariate Statistics

HSB

Variable	Descriptions or Sample Content of Items; Percentages for Categorical Items	Number of Items	Reliability	Mean	Standard Deviation
Achievement					
Vocabulary 1	Multiple-choice test	15	—	49.52	10.05
Vocabulary 2	Multiple-choice test	12	—	49.56	10.06
Reading	Multiple-choice test	20	—	49.37	10.12
Mathematics 1	Multiple-choice test	25	—	49.38	10.13
Mathematics 2	Multiple-choice test	8	—	49.85	9.96
Age	In years	1	—	17.47	8.46
Motivation					
Aspiration	Occupational goals age 30	1	—	8.47	4.39
Work orientation	(HSB scored composite)	3	—	.01	.89
	Self-satisfaction	2	.33	.90	.24
Adjustment Problems	Discipline problems, cutting, and suspension	3	.40	5.30	.80
Control locus	(HSB scored composite)	4	—	.01	.89
Motivation	Items on English and mathematics interest and usefulness, interest and hard work	11	.82	5.84	1.15
Law trouble	Serious trouble; Yes = 3.9%	1	—	1.96	.19
Quality of Instruction					
Quality	Ratings of quality of instruction such as good teaching, academic emphasis, school reputation, teacher interest in students, and instructional qualities	10	.59	5.19	.84
Quantity of Instruction					
Quantity	Academic courses completed in English, mathematics, French, German, Spanish, history and science	10	.71	4.84	.92
School Environment					
Facilities	Ratings of school building and library	2	.60	5.51	1.43
Discipline	Ratings of effectiveness and fairness of school	2	.69	4.63	1.50
Extracurricular activities	Student participation in school sports, clubs, band, and debate	15	.87	19.29	3.51
Peer Environment					
Peer	Grades of friends, their school interest in classes and college, and regular school attendance	4	.67	7.08	1.15
Home Environment					
Parent interest	Parental monitoring and interest in school, work, and career plans	9	.34	3.48	.81
Home facilities	Place to study, daily newspaper, enclyclopedia, and electric dishwasher	8	.52	14.01	1.53
Mother work	Mother working before and during elementary and high school	3	.78	6.12	2.02
Homework	Hours per week spent on homework	1	—	4.42	1.41
Age first worked	For pay	1	—	4.52	2.31
Worked last week	In hours	1	—	1.38	.48
Hours currently worked per week	In hours			3.77	1.66
Hours worked per week during previous school year	In hours			4.80	1.92
Socioeconomic status (SES)	SES composite scale (HSB scored)	5	—	−.09	.75
Media Exposure					
Television	Hours watched per day	1	—	4.43	1.70
Miscellaneous					
Handicaps	Visual, hearing, speech, learning, and health handicaps	8	.32	8.70	.60
Physically unattractive	Yes = 11.8	1	—	1.88	.31
Male	Yes = 47.9	1	—	1.48	.49
White	Yes = 75.5	1	—	.75	.43
Spanish	Yes = 11.2	1	—	.11	.32
Asian	Yes = 1.3	1	—	.01	.11
Black	Yes = 14.0	1	—	.14	.35
Alternative public schools	Yes = 3.2	1	—	.03	.18
Catholic	Yes = 9.5	1	—	.10	.29
Elite private	Yes = 1.1	1	—	.01	.11
Other private	Yes = 2.0	1	—	.02	.14

Table 1

Variable Description for IEA

Variable	Operational Definition, Scoring, and Sample Items
Mathematics Achievement	70 items assessing student mathematics in both content (mathematics subject matter) and cognitive-process levels (e.g., particular skills, abilities, and knowledge, etc.) Range of internal-consistency reliability (Kuder-Richardson 20) among 12 countries is.73 to .93, where the United States is .84. "The result of an operation on the numbers 9 and 18 is 27. In this operation, the number 27 is (a) product; (b) sum; (c) quotient; (d) difference, (e) average." "Four persons whose names begin with different letters are placed in a row, side by side. What is the probability that they will be placed in alphabetical order from left to right? (a) 1/120; (b) 1/24 (correct); (c) 1/12; (d) 1/6; (e) 1/4."
Male	Coded: 1-male, 0-female
Socioeconomic Status	Highest amount of either parent's education; code is indicated by the number of completed years.
Scientific Background	To indicate father's occupation: 1-scientific, 0-nonscientific
Father's Occupation	Coded: 1-higher professional and technical occupations 2-farm proprietors and farm laborers 3-subprofessional technical, small worker-proprietor (non-farm), clerical, and sales 4-manual workers (non-farm)
Mother's Employment	"Is your mother presently working?" Coded: 0-not working, 1-part time, 2-full time
Highest Mathematics Course Taken	"Indicate the highest level of mathematics courses that you have taken recently: (coded:) 1-arithmetic or general mathematics 2-algebra 3-geometry 4-trigonometry 5-advanced mathetics (calculus, etc.)"

Variable	Operational Definition, Scoring, and Sample Items
Number of Students in Math Class	"Indicate the approximate number of students in your present or most recent mathematics course: (coded:) 1-under 10 2-10 through 14 3-15 through 19 4-20 through 24 5-25 through 29 6-30 through 34 7-35 through 39 8-40 or more"
Periods of Math per Week	"In your mathematics class, how many periods do you have each week? (coded): 1-1 or 2 2-3 or 4 3-5 or 6 4-7 or 8 5-9 or 10 6-11 or 12 7-13 or more"
Hours of Homework	"Indicate the amount of hours that students usually devote to homework each week."
Extra Mathematics Activities	"Have you been a member of any mathematics club, or attended special lectures or courses on mathematics? (coded:) 1-yes, 0-no"
Interest in Mathematics	Ten items are included to measure the level of interest in mathematics; e.g., "Wishes to take additional math couses (coded:) 1-yes, 0-no" "Which two school subjects have you liked most? (coded:) 1-mathematics, 0-others
Attitude Toward	Eleven items form a scale to ascertain the student's disposition toward school life, e.g., "I find school interesting and challenging. (coded:) 2-agree, 0-disagree" "I am bored most of the time in school. (coded:) 2-disagree, 0-agree"